UNITED STATES PATENT APPLICATION FOR

A SYSTEM AND METHOD FOR ROBUST FOREGROUND AND BACKGROUND IMAGE DATA SEPARATION FOR LOCATION OF OBJECTS IN FRONT OF A CONTROLLABLE DISPLAY WITHIN A CAMERA VIEW

Inventors:

I-JONG LIN

NELSON LIANG AN CHANG

"Express Mail" mailing label number __EL 629876872 US__

Date of Deposit _____January 31, 2001_____

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, DC 20231.

_____Dee Timmons_____

(Typed or printed name of person mailing paper or fee)

_____
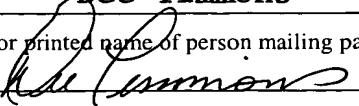
(Signature of person mailing paper or fee)

FIELD OF THE INVENTION

The present invention relates to a computer controllable display system and in particular to the interaction of a user with a computer controlled displayed image.

BACKGROUND OF THE INVENTION

Computer controlled projection systems generally include a computer system for generating image data and a projector for projecting the image data onto a projection screen. Typically, the computer controlled projection system is used to allow a presenter to project presentations that were created with the computer system onto a larger screen so that more than one viewer can easily see the presentation. Often, the presenter interacts with the projected image by pointing to notable areas on the projected image with his/her finger, laser pointer, or some other pointing device or instrument.

The problem with this type of system is that if a user wants to cause any change to the projected image, he/she must interact with the computer system using an input device such as a mouse, keyboard or remote device. For instance, a device is often employed by a presenter to remotely control the computer system via infrared signals to display the next slide in a presentation. However, this can be distracting to the viewers of the presentation since the presenter is no longer interacting with them and the projected presentation and, instead, is interacting with the computer system. Often, this interaction can lead to significant interruptions in the presentation.

Hence, a variation of the above system developed to overcome the computer-only interaction problem allows the presenter to directly interact with the projected image and thus better interaction with the audience. In this system, the computer generates image data (e.g. presentation slides) to be projected onto a projection screen with an image projector. The system also includes a digital image capture device such as a digital camera for capturing the projected image. The captured projected image data is transmitted back to the computing system and is used to determine the location of any objects (e.g., pointing device) in front of the screen. The computer system may then be controlled dependent on the determined location of the pointing device. For

Docket No.:10007845-1

instance, in U.S. Patent No. 5,138,304 assigned to the assignee of the subject

application, a light beam is projected onto the screen and is detected by a camera. To

determine the position of the light beam, the captured image data of the projected

image and the original image data are compared. The computer is then caused to

5   position a cursor in the video image at the pointer position or is caused to modify the

projected image data in response to the pointer position.

In order to implement a user interactive, computer controlled display or

projection system, it must be initially calibrated so as to determine the location of the

screen (i.e., the area in which the image is displayed) within the capture area of the

10   camera. Once the location of the screen is determined, this information can be used to

identify objects within the capture area that are within the display area but are not part

of the displayed image (e.g., objects in front of the display area). For instance, the

system can identify a pointer or finger in front of the display area and its location

within the display area. Knowing where objects are located in front of the display

15   area can be used to cause the system to respond to the object dependent on its location

within the display area.

In one known technique described in U.S. Patent No. 5,940,139, the

foreground and the background of a video are separated by illuminating the

foreground with a visible light and the background with a combination of infrared and

20   visible light and using two different cameras to pick of the signal and extract the

background from the foreground. In another known technique described in U.S.

Patent No. 5,345,308, a man-made object is discriminated within a video signal by

using a polarizer mounted to a video camera. The man-made object has both vertical

and horizontal surfaces that reflect light that can be polarized whereas, backgrounds

25   do not have polarizing components. Thus, the man-made object is filtered from the

video signal. These techniques are cumbersome in that they require additional

illumination methods, different types of cameras or filtering hardware and thus are not

conducive to exact object location or real-time operation in slide presentation

applications.

Docket No.:10007845-1

In still another known technique described in U.S. Patent No. 5,835,078, an infrared pointer is projected on a large screen display device, and the identity and location of the infrared pointer are determined. Specialized infrared pointing devices emit frequencies unique to each device. The identity and location of a given pointer is detected by detecting its frequency using an infrared camera. The identity and location of the pointer are then used to cause the computer system to display a mark corresponding to the given pointer on the large screen display at the point at which the infrared pointer is positioned. Although this technique identifies the location of an object projected on a display screen, it requires the use of specialized equipment including infrared pointers and infrared cameras. Moreover it relies upon the simple process of detecting infrared light on a displayed image. In contrast, the separation of a physical object in the foreground of a displayed image requires the actual separation of image data corresponding to the object from image data corresponding to the background of the object (i.e., foreground and background image separation).

The present invention is a technique for separating foreground and background image data of a display area within the capture area of an image capture device in a user interactive, computer controlled display system.

SUMMARY OF THE INVENTION

A system and method of locating objects positioned in front of a user interactive, computer controlled display area includes a computer system for displaying an image in the display area, means for converting the displayed image data into expected captured display area data using a derived coordinate location mapping function and a derived intensity mapping function, an image capture device for capturing the image in an image capture area to obtain captured data that includes captured display area data corresponding to a predetermined location of the display area in the capture area, and means for comparing the expected captured display area data to the captured display area data at each coordinate location of the captured display area data, such that non-matching compared image data corresponds to pixel locations of objects in front of the display area.

Docket No.:10007845-1

In another embodiment of the system including a computer controlled display area, the system is calibrated by displaying a plurality of calibration images within the display area each including a calibration object, capturing a plurality of images within the capture area each including one of the plurality of calibration images, determining

5   a mapping between the coordinate location of the calibration object in the display area and the coordinate location of the calibration object in the capture area for each captured image, and deriving a coordinate location mapping function from the location mappings of the plurality of captured images.

In another embodiment, the system is further calibrated by displaying at least

10   two intensity calibration objects having different displayed intensity values within the display area, capturing the intensity calibration objects within the capture area to obtain captured intensity values corresponding to the displayed intensity values, mapping the displayed intensity values to the captured intensity values, and deriving an intensity mapping function from the mappings between the displayed and captured

15   intensity values.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 illustrates a block diagram of a first embodiment of a system for locating objects in front of a display area in a user interactive, computer controlled

20   display system in accordance with the present invention;

Fig. 2A illustrates a first embodiment of the method of locating objects in front of a display area within a capture area in a user interactive, computer controlled display system in accordance with the present invention;

Fig. 2B illustrates converting display area image data into expected captured

25   display area image data;

Fig. 2C illustrates identifying captured display area image data using predetermined display area location information;

Docket No.:10007845-1

Fig. 2D illustrates comparing expected captured display area image data to captured display area image data;

Fig. 3 shows a capture area including an image of a display area and a hand positioned in front of the display area;

5   **Fig. 4** shows image data showing the location of the hand in the capture area illustrated in Fig. 3 obtained by performing the method illustrated in Fig. 2A in accordance with the present invention;

Fig. 5A illustrates a method of deriving a coordinate location function in accordance with the present invention;

10   **Fig. 5B** illustrates a calibration image including a calibration object;

Fig. 5C illustrates mapping the coordinate location of the calibration object in the displayed image coordinate system to the coordinate system of the captured displayed image; and

Fig. 6 shows a method of deriving an intensity mapping function in 15   accordance with the present invention.

DETAILED DESCRIPTION OF THE INVENTION

A block diagram of a user interactive, computer controlled image display system is shown in Fig. 1 including a computing system 10 for generating image data 10A and a graphical interface 11 for causing images 10B corresponding to the image data 10A to be displayed in display area 12. It should be understood that the graphical interface may be a portion of the computing system or may be a distinct element external to the computing system. The system further includes an image capture device 13 having an associated image capture area 13A for capturing displayed images 10B. The captured images also include images 10C of objects or regions that are outside of the display area 10B. The captured images can also include objects 10D that are positioned within the image capture area 13A in front of the display area 12. Non-display area images include anything other than what is displayed within the display area in response to image data 10A, including objects that extend into the display area. The captured images are converted into digital image data 13B and are transmitted to an object locator 14. Object locator 14 includes an image data converter 15 and an image data compare unit 16. The image data converter 15 converts display area image data 10A generated by the computing system into expected captured display area image data 15A using a derived coordinate location function and an intensity mapping function 15B. The expected image data 15A are coupled to image data compare unit 16 along with captured image data 13B and predetermined display area location information 13C. The image data compare unit 16 compares the expected captured display area image data 15A to the portion of the captured image data 13B that corresponds to the display area in the predetermined display area location. Non-matching compared data corresponds to the pixel locations in the captured display area image data 13B where an object is located. The object location information 16A can be transmitted to the computing system 10 for use in the user interactive, computer controlled display system.

In this embodiment, the computing system 10 includes at least a central processing unit (CPU) and a memory for storing digital data (e.g., image data) and has the capability of generating at least three levels of grayscale images. The display area

Docket No.:10007845-1

can be a computer monitor driven by the graphical interface or can be an area on a projection screen or projection area (e.g., a wall). In the case in which images are displayed using projection, the system includes an image projector (not shown in Fig. 1) that is responsive to image data provided from the graphical interface.

5   In one embodiment, the image capture device is a digital still or video camera or digital video camera arranged so as to capture at least all of the images 10B displayed in the display area 12 within a known time delay. It is well known in the field of digital image capture that an image is captured by a digital camera using an array of sensors that detect the intensity of the light impinging on the sensors within
10   the capture area of the camera. The light intensity signals are then converted into digital image data corresponding to the captured image. Hence, the captured image data 13B is digital image data corresponding to the captured image. In another embodiment the image capture device is an analog still or video camera and captured analog image data is converted into captured digital image data 13B.

15   In one embodiment, the images 10B correspond to a plurality of slides in a user's computer generated slide presentation.

        It should be noted that a single conversion of the displayed image data into expected captured image data is required per displayed image. However, more than one comparison can be performed per displayed image so as to detect the movement
20   and location of non-static objects positioned in front of the displayed image. For instance, while a single image is displayed it can be captured by image capture device 13 on a continual basis and each new captured image can be compared by image data compare unit 16 to the expected captured image data to locate objects at different time intervals.

25   It should be understood that all or a portion of the functions of the object locator 14 can be performed by the computing system. Consequently, although it is shown external to the computing system, all or portions of the object locator 14 may be implemented within the computing system.

Docket No.:10007845-1

It should be further understood that the object locator can be implemented in a software implementation, hardware implementation, or any combination of software and hardware implementations.

A first embodiment of a method for locating objects positioned in front of the display area 12 is shown in Fig. 2A. An image is displayed in the display area (block 20). The image can correspond to a current one of a plurality of images of a user's slide presentation being displayed during real-time use of the system shown in Fig. 1. It should be noted that the method as shown in Fig. 2A can be performed on each of the plurality of images (i.e., slides) of a slide presentation allowing the location of objects in front of the display area to be performed in real-time during the presentation.

The corresponding image data 10A (Fig. 1) employed by the computing system to display the image in the display area is converted into an expected captured display area data (block 21). The image data is converted using a derived coordinate location mapping function and a derived intensity mapping function. Fig. 2B illustrates the conversion of the display area image data to expected captured display area image data. The display area image 25 corresponds to the image data 10A generated by the computing system for either projecting or displaying an image. The image data 10A is converted using the derived coordinate location mapping function and intensity mapping function to generate data corresponding to the expected captured display area image 26.

The displayed image is captured in the capture area of an image capture device to obtain capture area image data (block 22). Fig. 2C shows the captured image data 27 that includes display area data 28 and non-display area image data 29. The display area data includes a portion of at least one object 30 that is located in front of the displayed image in the display area. As a result, the display area data includes image data corresponding to the portion of the object.

The location of the display area within the capture area is predetermined. This pre-determination can be performed during calibration of the system prior to real-time use of the user interactive, computer controlled display system. In one embodiment,

Docket No.:10007845-1

the pre-determination of the location of the display area is performed according to the system and method as disclosed in Application Serial No. _ (Attorney Docket No.: 10007846) incorporated herein by reference. Specifically, according to this method the location of the display area is determined by deriving constructive and destructive

5   feedback data from image data corresponding to a plurality of captured calibration images. It should be understood that other methods of determining the location of the display area in the capture area can be used to perform the system and method of locating objects in front of a display screen in accordance with the present invention. The pre-determination of the location of the display screen in the capture area allows

10   for the separation/identification of the captured display area data 31 from the captured image data 27 (Fig. 2C). In particular, as shown in Fig. 2C, the pre-determination of the location of the display area within the captured area allows for the separation/identification of only the display area data including both the displayed image data 28A and the data 28B corresponding to the portion of the object in front of

15   the display area.

The expected captured display area data 26 is compared to the identified captured display area data 31 by comparing mapped pixel values (block 23, Fig. 2D). Non-matching pixel values indicate the location of the object in front of the display area (block 24). As shown in Fig. 2D, the object 28B represents non-matching pixel

20   data thereby indicating an object in front of the display area.

It should be understood that although only a single conversion (block 21) of the displayed image data into expected captured image data is minimally required per displayed image, more than one comparison (block 23) can be performed per displayed image so as to detect the movement and location of non-static objects

25   positioned in front of the displayed image. For instance, while a single image is displayed it can be captured (block 22) on a continual basis and compared (block 23) to the expected captured image data to locate objects at during different time intervals as the image is being displayed.

Figs. 3 and 4 show images illustrating the method of locating objects in front

30   of a user interactive, computer controlled display system as shown in Fig. 2A. In

          Docket No.:10007845-1

particular, Fig. 3 shows the capture area 33 having an image including a display area 34 and an object 35 (i.e., a hand) positioned in front of the display area 34. Fig. 4 shows data obtained using the method shown in Fig. 2A to locate the hand in front of the display. In this example, the method of Fig. 2A additionally modifies the captured image data to show the location of the hand in front of the display area within the capture area by setting the pixel values (i.e., intensity values) at the coordinate locations 40 of the hand to one intensity value (e.g., white) and pixel values at the coordinate locations 41 where no objects are detected to a different intensity value (e.g., black).

In accordance with the method shown in Fig. 2A, captured display area data can be compared to expected display area data by subtracting the expected captured display area data (expected data) from the captured display area data (actual data) to obtain a difference value:

$$\delta(u_i, v_i) = \|ExpectedData(u_i, v_i) - ActualData(u_i, v_i)\| \qquad \text{Eq. 1}$$

where (u_i, v_i) are the coordinate locations in the capture display area. The difference value $\delta(u_i, v_i)$ is then compared to a threshold value, c_{thresh}, where c_{thresh} is a constant determined by the lighting conditions, image that is displayed, and camera quality. If the difference value is greater than the threshold value (i.e., $\delta(u_i, v_i) > c_{thresh}$) then an object exists at that coordinate point. In other words, the points on the display that do not meet the computer's intensity expected value at a given display area location have an object in the line of sight between the camera and the display.

Fig. 5A shows a method of calibrating a system for locating objects positioned in front of a user interactive, computer controlled display area. Calibration is achieved by initially displaying a plurality of coordinate calibration images (block 50). Fig. 5B shows an example of a coordinate calibration image 55 that includes a calibration object 54. The calibration images are characterized in that the calibration object is located at a different location within each of the calibration images. It should be noted that the object does not have to be circular in shape and can take other shapes to implement the method of the subject application.

Docket No.:10007845-1

The plurality of calibration images is successively captured in the capture area such that each captured image includes one of the calibration objects (block 51). For each captured image, the coordinate location of the display area calibration object is mapped to a coordinate location of the calibration object in the predetermined location of the display area in the capture area (block 52). It should be noted that the coordinate location of the display area calibration object is known from image data 10A (Fig. 1) and the coordinate location of the calibration object in the capture area is known from capture data 13B.

As shown in Fig. 5C, the displayed calibration image 55 can be viewed as having an x-y coordinate system and the captured image 58 can be viewed as having a u-v coordinate system, thus allowing the mapping of an x-y coordinate location of the calibration object 54 to a u-v coordinate location of the captured object 54'.

The image data corresponding to the display area 57 in the capture area is identified by predetermining the location of the display area within the capture area. As described above, display area location pre-determination can be performed according to the system and method as disclosed in Application Serial No. _____ (Attorney Docket No.: 10007846) however other methods can be used. The pre-determination of the location of the display screen in the capture area allows for the identification of the captured display area data and hence the mapping of the x-y coordinate location of the displayed calibration object 54 to a u-v coordinate location of the captured calibration object 54' in the predetermined display area.

The individual mappings of calibration object locations allow for the derivation of a function between the two coordinate systems (block 53):

$$(x, y) \xrightarrow{\ f\ } (u, v) \hspace{3cm} \text{Eq.2}$$

In one embodiment, a perspective transformation function (Eqs. 3 and 4) is used to derive the location mapping function:

$$f_u(x, y) = u = \frac{a_{11}x + a_{21}y + a_{31}}{a_{13}x + a_{23}y + a_{33}} \hspace{2cm} \text{Eq. 3}$$

Docket No.:10007845-1

$$f_v(x, y) = v = \frac{a_{12}x + a_{22}y + a_{32}}{a_{13}x + a_{23}y + a_{33}}$$

<div align="right">Eq. 4</div>

The variables a_{ij} of Eqs. 3 and 4 are derived by determining individual location mappings for each calibration object. It should be noted that other transformation functions can be used such as a simple translational mapping function or an affine

5   mapping function.

For instance, for a given calibration object in a calibration image displayed within the display area, its corresponding x,y coordinates are known from the image data 10A generated by the computer system. In addition, the u,v coordinates of the same calibration object in the captured calibration image are also known from the

10   portion of the captured image data 13B corresponding to the predetermined location of the display area in the capture area. The known x,y,u,v coordinate values are substituted into Eqs. 3 and 4 for the given calibration object. Each of the calibration objects in the plurality of calibration images are mapped in the same manner to obtain x and y calibration mapping equations (Eq. 3 and 4).

15   The location mappings of each calibration object are then used to derive the coordinate location functions (Eq. 3 and 4). Specifically, the calibration mapping equations are simultaneously solved to determine coefficients a_{11}-a_{33} of transformation functions Eqs. 3 and 4. Once determined, the coefficients are substituted into Eqs. 3 and 4 such that for any given x,y coordinate location in the

20   display area, a corresponding u-v coordinate location can be determined. It should be noted that an inverse mapping function from u-v coordinates to x,y coordinates can also be derived from the coefficients a_{11}-a_{33}.

In the case of a two-dimensional transformation function (e.g., Eqs. 3 and 4), nine coefficients (e.g., a_{11}-a_{33}) need to be determined and, hence at least nine

25   equations are required. Since, there are two mapping equations per calibration image, at least five calibration images are required in order to solve for the function. It should be noted that more calibration objects may be used and this overconstrained problem (i.e., more calibration objects than required to solve for the coefficients) may be robustly approximated with LSQ (i.e., least square) fit.

The method shown in Fig. 5A can further include the calibration method shown in Fig. 6 for determining an intensity mapping function. Calibration is achieved by displaying at least two intensity calibration objects having different intensity values from the other (block 60). The, at least, two intensity calibration objects may be displayed in separate images or with the same image. The, at least, two objects may be displayed at the same location or different locations within the image or images. The intensity calibration objects can be a color or a grayscale image object. The displayed intensity values of the displayed intensity calibration objects are known from the image data 10A generated by the computing system 10 (Fig. 1). The, at least, two calibration objects are captured (block 61) to obtain capture data 13B where the captured objects have associated captured intensity values corresponding to the displayed intensity values. The displayed intensity values are mapped to the captured intensity values (block 62). An intensity mapping function is derived from the, at least, two intensity mappings (block 63). It should be noted that the derived coordinate location mapping function is used to identify corresponding pixel locations between the display area and the captured display area to allow for intensity mapping between pixels at the corresponding locations.

In one embodiment, the intensity mapping function is determined using interpolation. For example, given the mappings between the displayed and captured intensity values, a range of displayed values and corresponding mapped captured values can be determined using linear interpolation. Captured and interpolated captured intensity values can then be stored in a look-up table such that when a displayed intensity value accesses the table, a corresponding mapped captured intensity value can be obtained. It should be noted that the mapping is not limited to linear interpolation and other higher order or non-linear interpolation methods can be employed.

Hence, the intensity and coordinate location mapping functions are determined so as to calculate $ExpectedData(u_i, v_i)$ in Eq. 1 . The absolute difference (i.e., $\delta(u_i, v_i)$) between the $ExpectedData(u_i, v_i)$ and $ActualData(u_i, v_i)$ is then determined to locate the object in the display area of the captured data.

Docket No.:10007845-1

A system and method is described that provides an arithmetically non-complex solution to locating objects in front of a display area within the capture area of an image capture device in a user interactive, computer controlled display system. Specifically, a system is described whereas an image is displayed on a per frame basis

5 and a simple series of operations are performed continuously to determine the location of the object(s) in front of the displayed image.

In the preceding description, numerous specific details are set forth, such as calibration image type and a perspective transformation function in order to provide a thorough understanding of the present invention. It will be apparent, however, to one

10 skilled in the art that these specific details need not be employed to practice the present invention. In other instances, well-known image processing techniques have not been described in detail in order to avoid unnecessarily obscuring the present invention.

In addition, although elements of the present invention have been described in

15 conjunction with certain embodiments, it is appreciated that the invention can be implement in a variety of other ways. Consequently, it is to be understood that the particular embodiments shown and described by way of illustration is in no way intended to be considered limiting. Reference to the details of these embodiments is not intended to limit the scope of the claims which themselves recited only those

20 features regarded as essential to the invention.

Docket No.:10007845-1